

## IN THE CLAIMS

Claims 1, 3-5, 7-9, 11-14, 16, 17, 20-33, 35, 68-71 and 73-80 are pending.

Claims 2, 6, 10, 15, 18, 19, 34, 36-67 and 72 were previously canceled.

Claims 76-80 were previously added.

Claims 1, 3, 4, 21-25, 31, 32, 68, 70, 74 and 75 are currently amended.

1. (Currently Amended) A method implemented at least in part by a computing device of identifying one or more portions of a document described by a tree structure having a plurality of nodes, the method comprising:

identifying a plurality of visual blocks in the document based on, at least, a document model of the document;

detecting, distinct from the plurality of visual blocks, one or more separators of the document based on, at least, one or more characteristics of at least one of the plurality of visual blocks; ~~and~~

assigning, to each of the one or more separators, a weight based on characteristics of visual blocks on either side of the separator; and

constructing, based at least in part on the plurality of visual blocks and the one or more separators, a content structure for the document, wherein the content structure identifies the different visual blocks as different portions of semantic content of the document.

2. (Canceled)

3. **(Currently Amended)** A method as recited in claim 1, ~~wherein the document is described by a tree structure having a plurality of nodes, and~~ wherein identifying the plurality of visual blocks in the document comprises:
- identifying a group of candidate nodes of the plurality of nodes;
  - for the respective nodes in the group of candidate nodes:
    - determining whether the node can be divided, and
    - if the node cannot be divided, then identifying the node as representing a visual block.
4. **(Currently Amended)** A method as recited in claim 3, wherein if the node cannot be divided, then based on a plurality of rules, setting a degree of coherence for the visual block represented by the node.
5. **(Original)** A method as recited in claim 3, wherein if the node cannot be divided, then removing the node from the group of candidate nodes.
6. **(Canceled)**
7. **(Original)** A method as recited in claim 3, wherein determining whether the node can be divided comprises determining that the node can be divided if

a background color of the node is different from a background color of a child of the node.

8. (Original) A method as recited in claim 3, further comprising checking whether the node has a child having a width and height greater than zero, and if the node has no child having a width and height greater than zero then removing the node from the group of candidate nodes.
9. (Original) A method as recited in claim 3, wherein determining whether the node can be divided comprises determining that the node can be divided if a size of the node is at least a threshold amount greater than a sum of sizes of children nodes of the node.
10. (Canceled)
11. (Original) A method as recited in claim 1, wherein the document is described by a tree structure having a plurality of nodes, and wherein identifying the plurality of visual blocks in the document comprises identifying different visual blocks based at least in part on HyperText Markup Language (HTML) tags of the plurality of nodes.

12. (Original) A method as recited in claim 1, wherein the document is described by a tree structure having a plurality of nodes, and wherein identifying the plurality of visual blocks in the document comprises identifying different visual blocks based at least in part on background colors of the plurality of nodes.

13. (Original) A method as recited in claim 1, wherein the document is described by a tree structure having a plurality of nodes, and wherein identifying the plurality of visual blocks in the document comprises identifying different visual blocks based at least in part on whether the plurality of nodes include text and the sizes of the plurality of nodes.

14. (Previously Presented) A method as recited in claim 1, wherein the document has, at least, a horizontal direction and a vertical direction;  
and

detecting the one or more separators comprises:

detecting one or more horizontal separators of the document;

and

detecting one or more vertical separators of the document.

15. (Canceled)

16. (Previously Presented) A method as recited in claim 1, further comprising determining to split a particular one of the separators into multiple separators if one or more of the plurality of visual blocks is contained in the particular separator.

17. (Previously Presented) A method as recited in claim 1, further comprising determining, if one or more of the plurality of visual blocks overlap a particular one of the separators, to modify one or more parameters of the particular separator so that the one or more of the plurality of visual blocks no longer overlap the particular separator.

18. (Canceled)

19. (Canceled)

20. (Previously Presented) A method as recited in claim 1, further comprising determining to remove a particular one of the separators from a separator list if one or more of the plurality of visual blocks cover the particular separator.

21. **(Currently Amended)** A method as recited in claim 1, ~~further comprising assigning, to each of the one or more separators, a weight based on~~

~~characteristics of visual blocks on either side of the separator, wherein the weight of the separator indicates how visible the separator is when the document is displayed.~~

22. **(Currently Amended)** A method as recited in claim 1 ~~21~~, wherein assigning the weight comprises assigning the weight based on a distance between two visual blocks on either side of the separator.
23. **(Currently Amended)** A method as recited in claim 1 ~~21~~, wherein assigning the weight comprises assigning the weight based on whether the separator is at a same position as an <HR> HTML tag.
24. **(Currently Amended)** A method as recited in claim 1 ~~21~~, wherein assigning the weight comprises assigning the weight based on a font size used in two visual blocks on either side of the separator.
25. **(Currently Amended)** A method as recited in claim 1 ~~21~~, wherein assigning the weight comprises assigning the weight based on a background color used in two visual blocks on either side of the separator.
26. **(Original)** A method as recited in claim 1, further comprising:

checking whether each of the plurality of visual blocks satisfies a degree of coherence threshold; and

for each of the plurality of visual blocks that does not satisfy the degree of coherence threshold, identifying a new plurality of visual blocks in the visual block, and repeating the detecting and constructing using the new plurality of visual blocks.

27. (Original) A method as recited in claim 1, wherein constructing the content structure comprises:

generating one or more virtual blocks based on the plurality of visual blocks; and

including, in the content structure, the one or more virtual blocks.

28. (Original) A method as recited in claim 27, wherein generating the one or more virtual blocks comprises generating the one or more virtual blocks by combining two visual blocks of the plurality of visual blocks.

29. (Original) A method as recited in claim 27, further comprising:

determining a degree of coherence value for each of the one or more virtual blocks.

30. (Original) A method as recited in claim 29, wherein determining the degree of coherence value for a virtual block comprises determining the degree of coherence value for the virtual block based at least in part on a weight of a separator between two visual blocks used to generate the virtual block.

31. (Currently Amended) One or more computer readable media having stored thereon a plurality of instructions that, when executed by one or more processors of a device, causes the one or more processors to, at least:

identify visual blocks in a document based on, at least, a document model wherein the said document is described by a tree structure having a plurality of nodes;

detect, distinct from the visual blocks, visual separators of the document based on, at least, one or more characteristics of at least one of the visual blocks; ~~and~~

assign to each of the one or more separators, a weight based on characteristics of visual blocks on either side of the separator; and

construct, based at least in part on the visual blocks and the visual separators, a content structure for the document that identifies regions of the document that represent semantic content of the document.

32. (Currently Amended) One or more computer readable media as recited in claim 31, wherein ~~the document is described by a tree structure having a~~



~~plurality of nodes, and wherein~~ the instructions that cause the one or more processors to identify visual blocks in the document comprise instructions that cause the one or more processors to:

identify a group of candidate nodes of the plurality of nodes;

for each node in the group of candidate nodes:

determine whether the node can be divided, and

if the node cannot be divided, then identify the node as representing a visual block.

33. (Previously Presented) One or more computer readable media as recited in claim 31, wherein:

the document has, at least, a horizontal direction and a vertical direction;

and

the instructions that cause the one or more processors to detect visual separators comprise instructions that cause the one or more processors to, at least:

detect one or more horizontal separators of the document; and

detect one or more vertical separators of the document.

34. (Canceled)

35. (Original) One or more computer readable media as recited in claim 31, wherein the instructions further cause the one or more processors to:

to check whether each of the visual blocks satisfies a degree of coherence threshold; and

for each of the visual blocks that does not satisfy the degree of coherence threshold, identify new visual blocks in the visual block, and repeat the detection and construction using the new visual blocks.

36 - 67. (Canceled)

68. (Currently Amended) A system, implemented at least in part by a computing device, comprising:

a visual block extractor, embodied at least in part in a computer readable medium, to extract visual blocks from a document based on, at least, a document model wherein the said document is described by a tree structure having a plurality of nodes;

a visual separator detector, embodied at least in part in a computer readable medium, coupled to receive the extracted visual blocks and configured to, at least, detect, based on, at least, one or more characteristics of the extracted visual blocks, one or more visual separators of the document and assign to each of the one or more separators, a weight based on characteristics of visual blocks on either side of the separator; and

a content structure constructor, embodied at least in part in a computer readable medium, coupled to receive the extracted visual blocks and the detected visual separators, and configured to, at least, construct a content structure for the document based on, at least:

one or more of the extracted visual blocks; and

one or more of the visual separators.

69. (Original) A system as recited in claim 68, further comprising:

a document retrieval module to retrieve documents from a plurality of documents based at least in part on the content structure constructed for one or more of the plurality of documents.

70. (Currently Amended) A system as recited in claim 68, ~~wherein the document is described by a tree structure having a plurality of nodes, and~~ wherein the visual block extractor is to extract visual blocks from the document by:

identifying a group of candidate nodes of the plurality of nodes;

for each node in the group of candidate nodes:

determining whether the node can be divided, and

if the node cannot be divided, then identifying the node as representing a visual block.

71. (Previously Presented) A system as recited in claim 68, wherein:

the document has, at least, a horizontal direction and a vertical direction; and

the visual separator detector is further configured to, at least:

detect one or more horizontal separators of the document and;

detect one or more vertical separators of the document.

72. (Canceled)

73. (Original) A system as recited in claim 68, wherein the content structure constructor is further to:

check whether each of the plurality of visual blocks satisfies a degree of coherence threshold; and

for each of the plurality of visual blocks that does not satisfy the degree of coherence threshold, return the visual block to the visual block extractor to have a new plurality of visual blocks extracted from the visual block, and further to have the visual separator detector detect one or more visual separators using the new plurality of visual blocks.

74. **(Currently Amended)** A system, implemented at least in part by a computing device, comprising:

means, embodied at least in part in a computer readable medium, for identifying a plurality of visual blocks in a document based on, at least, a document model of the document wherein the said document is described by a tree structure having a plurality of nodes;

means, embodied at least in part in a computer readable medium, for detecting, distinct from the plurality of visual blocks, one or more separators of the document based on, at least, one or more characteristics of at least one of the plurality of visual blocks, and assigning to each of the one or more separators, a weight based on characteristics of visual blocks on either side of the separator; and

means, embodied at least in part in a computer readable medium, for constructing, based at least in part on the plurality of visual blocks and the one or more separators, a content structure for the document, wherein the content structure identifies the different visual blocks as different portions of semantic content of the document.

75. **(Currently Amended)** A system as recited in claim 74, ~~wherein the document is described by a tree structure having a plurality of nodes, and~~ wherein the means for identifying the plurality of visual blocks in the document comprises:

means, embodied at least in part in a computer readable medium, for identifying a group of candidate nodes of the plurality of nodes;

for each node in the group of candidate nodes:

means, embodied at least in part in a computer readable medium, for determining whether the node can be divided, and

means, embodied at least in part in a computer readable medium, for identifying, if the node cannot be divided, the node as representing a visual block.

76. (Previously Added) A method as recited in claim 1, wherein:

visual blocks are specified with respect to the document model; and

separators are specified with respect to the document as it would be displayed.

77. (Previously Added) A method as recited in claim 76, wherein the separator specification comprises a specification of a display area.

78. (Previously Added) A method as recited in claim 77, wherein the specification of the display area comprises a specification of a start pixel and a specification of an end pixel.

79. (Previously Added) A method as recited in claim 1, wherein detecting one or more separators of the document comprises initializing a specification of an initial separator to include a display area that would be occupied by the entire document if it were displayed.

80. (Previously Added) A method as recited in claim 1, wherein detecting one or more separators of the document comprises initializing a specification of an initial separator to include a display area that would contain each of the plurality of visual blocks if they were displayed.